

CL-1806-conversion-CIP 3-5-04.ST25
SEQUENCE LISTING

<110> E.I. DuPont de Nemours, & Company
Yadav, Narendra S.
Yang, Jianjun Gene

<120> Intein-Mediated Protein Splicing

<130> CL1806 US CIP

<150> US 60/354395

<151> 2002-02-04

<160> 78

<170> PatentIn version 3.2

<210> 1

<211> 123

<212> PRT

<213> Synechocystis sp. PCC6803

<400> 1

Cys Leu Ser Phe Gly Thr Glu Ile Leu Thr Val Glu Tyr Gly Pro Leu
1 5 10 15

Pro Ile Gly Lys Ile Val Ser Glu Glu Ile Asn Cys Ser Val Tyr Ser
20 25 30

Val Asp Pro Glu Gly Arg Val Tyr Thr Gln Ala Ile Ala Gln Trp His
35 40 45

Asp Arg Gly Glu Gln Glu Val Leu Glu Tyr Glu Leu Glu Asp Gly Ser
50 55 60

Val Ile Arg Ala Thr Ser Asp His Arg Phe Leu Thr Thr Asp Tyr Gln
65 70 75 80

Leu Leu Ala Ile Glu Glu Ile Phe Ala Arg Gln Leu Asp Leu Leu Thr
85 90 95

Leu Glu Asn Ile Lys Gln Thr Glu Glu Ala Leu Asp Asn His Arg Leu
100 105 110

Pro Phe Pro Leu Leu Asp Ala Gly Thr Ile Lys
115 120

<210> 2

<211> 37

<212> PRT

<213> Synechocystis sp. PCC6803

<400> 2

Met Val Lys Val Ile Gly Arg Arg Ser Leu Gly Val Gln Arg Ile Phe
1 5 10 15

Asp Ile Gly Leu Pro Gln Asp His Asn Phe Leu Leu Ala Asn Gly Ala
 20 25 30

Ile Ala Ala Asn Cys
 35

<210> 3
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 3
 tgcctttctt tcggaactga gatccttacc gttgagtacg gaccacttcc tattggtaag 60
 atcgtttctg aggaa 75

<210> 4
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 4
 attaactgct cagtgtactc tgttgatcca gaaggaagag ttacactca ggctatcgca 60
 caatggcacg atagg 75

<210> 5
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 5
 ggtgaacaag aggttctcga gtacgagctt gaagatggat ccgttattcg tgctacctct 60
 gaccatagat tcttg 75

<210> 6
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 6
 actacagatt atcagcttct cgctatcgag gaaatctttg ctaggcaact tgatctcctt 60
 actttggaga acatc 75

CL-1806-conversion-CIP 3-5-04.ST25

<210> 7
 <211> 69
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 7
 aagcagacag aagaggctct tgacaaccac agacttccat tccctttgct cgatgctgga 60
 accatcaag 69

<210> 8
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 8
 cttgatggtt ccagcatcga gcaaagggaa 30

<210> 9
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 9
 tggaagtctg tggttgtcaa gagcctcttc tgtctgcttg atgttctcca aagtaaggag 60
 atcaagttgc ctagc 75

<210> 10
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 10
 aaagatttcc tcgatagcga gaagctgata atctgtagtc aagaatctat ggtcagaggt 60
 agcacgaata acgga 75

<210> 11
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 11

tccatcttca agctcgctact cgagaacctc ttgttcaccc ctatcgtgcc attgtgcat 60
 agcctgagtg taaac 75

<210> 12
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 12
 tcttccttct ggatcaacag agtacactga gcagttaatt tcctcagaaa cgatcttacc 60
 aataggaagt ggtcc 75

<210> 13
 <211> 39
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 13
 gtactcaacg gtaaggatct cagttccgaa agaaaggca 39

<210> 14
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 14
 atgggtaagg tgattggaag acgttctctt ggtgttcaaa ggatcttcga tatcggattg 60
 ccacaagacc acaac 75

<210> 15
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 15
 tttcttctcg ctaatggtgc catcgctgcc aattgc 36

<210> 16
 <211> 75
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Modified from Synechocystis sp. PCC6803, to contain plant

preferred codons

<400> 16
gcaattggca gcgatggcac cattagcgag aagaaagttg tggctctgtg gcaatccgat 60
atcgaagatc ctttg 75

<210> 17
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 17
aacaccaaga gaacgtcttc caatcacctt aaccat 36

<210> 18
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 18
tgcctttctt tcggaactga g 21

<210> 19
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 19
tcacttgatg gttccagcat cgag 24

<210> 20
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 20
ccatggttaa ggtgattgga agac 24

<210> 21
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified from Synechocystis sp. PCC6803, to contain plant preferred codons

<400> 21
gcaattggca gcgatggcac c 21

<210> 22
<211> 369
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified from Synechocystis sp. PCC6803 SspE, to contain plant preferred codons

<220>
<221> CDS
<222> (1)..(369)
<223> Modified from Synechocystis sp. PCC6803 SspE, to contain plant preferred codons

<400> 22
tgc ctt tct ttc gga act gag atc ctt acc gtt gag tac gga cca ctt 48
Cys Leu Ser Phe Gly Thr Glu Ile Leu Thr Val Glu Tyr Gly Pro Leu
1 5 10 15

cct att ggt aag atc gtt tct gag gaa att aac tgc tca gtg tac tct 96
Pro Ile Gly Lys Ile Val Ser Glu Glu Ile Asn Cys Ser Val Tyr Ser
20 25 30

gtt gat cca gaa gga aga gtt tac act cag gct atc gca caa tgg cac 144
Val Asp Pro Glu Gly Arg Val Tyr Thr Gln Ala Ile Ala Gln Trp His
35 40 45

gat agg ggt gaa caa gag gtt ctc gag tac gag ctt gaa gat gga tcc 192
Asp Arg Gly Glu Gln Glu Val Leu Glu Tyr Glu Leu Glu Asp Gly Ser
50 55 60

gtt att cgt gct acc tct gac cat aga ttc ttg act aca gat tat cag 240
Val Ile Arg Ala Thr Ser Asp His Arg Phe Leu Thr Thr Asp Tyr Gln
65 70 75 80

ctt ctc gct atc gag gaa att ttt gct agg caa ctt gat ctc ctt act 288
Leu Leu Ala Ile Glu Glu Ile Phe Ala Arg Gln Leu Asp Leu Leu Thr
85 90 95

ttg gag aac att aag cag aca gaa gag gct ctt gac aac cac aga ctt 336
Leu Glu Asn Ile Lys Gln Thr Glu Glu Ala Leu Asp Asn His Arg Leu
100 105 110

cca ttc cct ttg ctc gat gct gga acc atc aag 369
Pro Phe Pro Leu Leu Asp Ala Gly Thr Ile Lys
115 120

<210> 23
<211> 123
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 23
Cys Leu Ser Phe Gly Thr Glu Ile Leu Thr Val Glu Tyr Gly Pro Leu
1 5 10 15

CL-1806-conversion-CIP 3-5-04.ST25

Pro Ile Gly Lys Ile Val Ser Glu Glu Ile Asn Cys Ser Val Tyr Ser
20 25 30

Val Asp Pro Glu Gly Arg Val Tyr Thr Gln Ala Ile Ala Gln Trp His
35 40 45

Asp Arg Gly Glu Gln Glu Val Leu Glu Tyr Glu Leu Glu Asp Gly Ser
50 55 60

Val Ile Arg Ala Thr Ser Asp His Arg Phe Leu Thr Thr Asp Tyr Gln
65 70 75 80

Leu Leu Ala Ile Glu Glu Ile Phe Ala Arg Gln Leu Asp Leu Leu Thr
85 90 95

Leu Glu Asn Ile Lys Gln Thr Glu Glu Ala Leu Asp Asn His Arg Leu
100 105 110

Pro Phe Pro Leu Leu Asp Ala Gly Thr Ile Lys
115 120

<210> 24
<211> 111
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified from Synechocystis sp. PCC6803 SspE, to contain plant preferred codons

<220>
<221> CDS
<222> (1)..(111)
<223> Modified from Synechocystis sp. PCC6803 SspE, to contain plant preferred codons

<400> 24
atg gtt aag gtg att gga aga cgt tct ctt ggt gtt caa agg atc ttc 48
Met Val Lys Val Ile Gly Arg Arg Ser Leu Gly Val Gln Arg Ile Phe
1 5 10 15

gat atc gga ttg cca caa gac cac aac ttt ctt ctc gct aat ggt gcc 96
Asp Ile Gly Leu Pro Gln Asp His Asn Phe Leu Leu Ala Asn Gly Ala
20 25 30

atc gct gca aat tgc 111
Ile Ala Ala Asn Cys
35

<210> 25
<211> 37
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 25

CL-1806-conversion-CIP 3-5-04.ST25

Met Val Lys Val Ile Gly Arg Arg Ser Leu Gly Val Gln Arg Ile Phe
1 5 10 15

Asp Ile Gly Leu Pro Gln Asp His Asn Phe Leu Leu Ala Asn Gly Ala
20 25 30

Ile Ala Ala Asn Cys
35

<210> 26
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer to introduce CDS for peptide MAHHHHHH at the N-terminus of GUS

<400> 26
atggctcatc atcatcatca tcatgtacgt cctgtagaaa cccaacc 48

<210> 27
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer to introduce a BamHI site after the stop codon of GUS

<400> 27
ggatccttgt ttgcctcct gctgcgg 27

<210> 28
<211> 618
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified GUS protein, with 6x His tags on the N-terminus and C-terminus

<220>
<221> PEPTIDE
<222> (1)..(618)
<223> Modified GUS protein, with 6x His tags on the N-terminus and C-terminus

<400> 28

Met Ala His His His His His Val Arg Pro Val Glu Thr Pro Thr
1 5 10 15

Arg Glu Ile Lys Lys Leu Asp Gly Leu Trp Ala Phe Ser Leu Asp Arg
20 25 30

Glu Asn Cys Gly Ile Asp Gln Arg Trp Trp Glu Ser Ala Leu Gln Glu
35 40 45

Ser Arg Ala Ile Ala Val Pro Gly Ser Phe Asn Asp Gln Phe Ala Asp
 50 55 60
 Ala Asp Ile Arg Asn Tyr Ala Gly Asn Val Trp Tyr Gln Arg Glu Val
 65 70 75 80
 Phe Ile Pro Lys Gly Trp Ala Gly Gln Arg Ile Val Leu Arg Phe Asp
 85 90 95
 Ala Val Thr His Tyr Gly Lys Val Trp Val Asn Asn Gln Glu Val Met
 100 105 110
 Glu His Gln Gly Gly Tyr Thr Pro Phe Glu Ala Asp Val Thr Pro Tyr
 115 120 125
 Val Ile Ala Gly Lys Ser Val Arg Ile Thr Val Cys Val Asn Asn Glu
 130 135 140
 Leu Asn Trp Gln Thr Ile Pro Pro Gly Met Val Ile Thr Asp Glu Asn
 145 150 155 160
 Gly Lys Lys Lys Gln Ser Tyr Phe His Asp Phe Phe Asn Tyr Ala Gly
 165 170 175
 Ile His Arg Ser Val Met Leu Tyr Thr Thr Pro Asn Thr Trp Val Asp
 180 185 190
 Asp Ile Thr Val Val Thr His Val Ala Gln Asp Cys Asn His Ala Ser
 195 200 205
 Val Asp Trp Gln Val Val Ala Asn Gly Asp Val Ser Val Glu Leu Arg
 210 215 220
 Asp Ala Asp Gln Gln Val Val Ala Thr Gly Gln Gly Thr Ser Gly Thr
 225 230 235 240
 Leu Gln Val Val Asn Pro His Leu Trp Gln Pro Gly Glu Gly Tyr Leu
 245 250 255
 Tyr Glu Leu Cys Val Thr Ala Lys Ser Gln Thr Glu Cys Asp Ile Tyr
 260 265 270
 Pro Leu Arg Val Gly Ile Arg Ser Val Ala Val Lys Gly Glu Gln Phe
 275 280 285
 Leu Ile Asn His Lys Pro Phe Tyr Phe Thr Gly Phe Gly Arg His Glu
 290 295 300
 Asp Ala Asp Leu Arg Gly Lys Gly Phe Asp Asn Val Leu Met Val His
 305 310 315 320

Asp His Ala Leu Met Asp Trp Ile Gly Ala Asn Ser Tyr Arg Thr Ser
 325 330 335
 His Tyr Pro Tyr Ala Glu Glu Met Leu Asp Trp Ala Asp Glu His Gly
 340 345 350
 Ile Val Val Ile Asp Glu Thr Ala Ala Val Gly Phe Asn Leu Ser Leu
 355 360 365
 Gly Ile Gly Phe Glu Ala Gly Asn Lys Pro Lys Glu Leu Tyr Ser Glu
 370 375 380
 Glu Ala Val Asn Gly Glu Thr Gln Gln Ala His Leu Gln Ala Ile Lys
 385 390 395 400
 Glu Leu Ile Ala Arg Asp Lys Asn His Pro Ser Val Val Met Trp Ser
 405 410 415
 Ile Ala Asn Glu Pro Asp Thr Arg Pro Gln Gly Ala Arg Glu Tyr Phe
 420 425 430
 Ala Pro Leu Ala Glu Ala Thr Arg Lys Leu Asp Pro Thr Arg Pro Ile
 435 440 445
 Thr Cys Val Asn Val Met Phe Cys Asp Ala His Thr Asp Thr Ile Ser
 450 455 460
 Asp Leu Phe Asp Val Leu Cys Leu Asn Arg Tyr Tyr Gly Trp Tyr Val
 465 470 475 480
 Gln Ser Gly Asp Leu Glu Thr Ala Glu Lys Val Leu Glu Lys Glu Leu
 485 490 495
 Leu Ala Trp Gln Glu Lys Leu His Gln Pro Ile Ile Ile Thr Glu Tyr
 500 505 510
 Gly Val Asp Thr Leu Ala Gly Leu His Ser Met Tyr Thr Asp Met Trp
 515 520 525
 Ser Glu Glu Tyr Gln Cys Ala Trp Leu Asp Met Tyr His Arg Val Phe
 530 535 540
 Asp Arg Val Ser Ala Val Val Gly Glu Gln Val Trp Asn Phe Ala Asp
 545 550 555 560
 Phe Ala Thr Ser Gln Gly Ile Leu Arg Val Gly Gly Asn Lys Lys Gly
 565 570 575
 Ile Phe Thr Arg Asp Arg Lys Pro Lys Ser Ala Ala Phe Leu Leu Gln
 580 585 590

Lys Arg Trp Thr Gly Met Asn Phe Gly Glu Lys Pro Gln Gln Gly Gly
 595 600 605

Lys Gln Gly Ser His His His His His His
 610 615

<210> 29
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer to amplify GUS

<400> 29
 cgcagcgtaa tgctctacac c 21

<210> 30
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer to amplify GUS

<400> 30
 ccgtaataac gggtcaggca c 21

<210> 31
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer to amplify the 2 um yeast replication origin and a Trp
 selective marker

<400> 31
 aggggaacaaa agctggagct ccaccagagg gccaaagaggg agggc 45

<210> 32
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer to amplify the 2 um yeast replication origin and a Trp
 selective marker

<400> 32
 cactagttct agagcggccg ccacatatg atccaatatc aaagg 45

<210> 33
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer for PCR-directed recombination for in-frame fusion of
 GUS-n/ Int-n and Int-c/GUS-c

<400> 33

ggatctcagt tccgaaagaa aggcagtctt gcgcgacatg cgtca 45

<210> 34
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer for PCR-directed recombination for in-frame fusion of
 GUS-n/ Int-n and Int-c/GUS-c

<400> 34
 cccctcgagg tcgacggat c gatatccat ggctcatcat catca 45

<210> 35
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer for PCR-directed recombination for in-frame fusion of
 GUS-n/ Int-n and Int-c/GUS-c

<400> 35
 gtccgtactc aacggtaagg atctcgtctt gcgcgacatg cgtca 45

<210> 36
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer for PCR-directed recombination for in-frame fusion of
 GUS-n/ Int-n and Int-c/GUS-c

<400> 36
 cgctaattggt gccatcgctg ccaattgtaa ccacgcgtct gttga 45

<210> 37
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer for PCR-directed recombination for in-frame fusion of
 GUS-n/ Int-n and Int-c/GUS-c

<400> 37
 cgaggtcgac ggtatcgata ag 22

<210> 38
 <211> 326
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> GUSn/Intn fusion

<220>
 <221> PEPTIDE
 <222> (1)..(326)
 <223> GUSn/Intn fusion

CL-1806-conversion-CIP 3-5-04.ST25

<400> 38

Met Ala His His His His His Val Arg Pro Val Glu Thr Pro Thr
1 5 10 15

Arg Glu Ile Lys Lys Leu Asp Gly Leu Trp Ala Phe Ser Leu Asp Arg
20 25 30

Glu Asn Cys Gly Ile Asp Gln Arg Trp Trp Glu Ser Ala Leu Gln Glu
35 40 45

Ser Arg Ala Ile Ala Val Pro Gly Ser Phe Asn Asp Gln Phe Ala Asp
50 55 60

Ala Asp Ile Arg Asn Tyr Ala Gly Asn Val Trp Tyr Gln Arg Glu Val
65 70 75 80

Phe Ile Pro Lys Gly Trp Ala Gly Gln Arg Ile Val Leu Arg Phe Asp
85 90 95

Ala Val Thr His Tyr Gly Lys Val Trp Val Asn Asn Gln Glu Val Met
100 105 110

Glu His Gln Gly Gly Tyr Thr Pro Phe Glu Ala Asp Val Thr Pro Tyr
115 120 125

Val Ile Ala Gly Lys Ser Val Arg Ile Thr Val Cys Val Asn Asn Glu
130 135 140

Leu Asn Trp Gln Thr Ile Pro Pro Gly Met Val Ile Thr Asp Glu Asn
145 150 155 160

Gly Lys Lys Lys Gln Ser Tyr Phe His Asp Phe Phe Asn Tyr Ala Gly
165 170 175

Ile His Arg Ser Val Met Leu Tyr Thr Thr Pro Asn Thr Trp Val Asp
180 185 190

Asp Ile Thr Val Val Thr His Val Ala Gln Asp Cys Leu Ser Phe Gly
195 200 205

Thr Glu Ile Leu Thr Val Glu Tyr Gly Pro Leu Pro Ile Gly Lys Ile
210 215 220

Val Ser Glu Glu Ile Asn Cys Ser Val Tyr Ser Val Asp Pro Glu Gly
225 230 235 240

Arg Val Tyr Thr Gln Ala Ile Ala Gln Trp His Asp Arg Gly Glu Gln
245 250 255

Glu Val Leu Glu Tyr Glu Leu Glu Asp Gly Ser Val Ile Arg Ala Thr
Page 13

260

265

270

Ser Asp His Arg Phe Leu Thr Thr Asp Tyr Gln Leu Leu Ala Ile Glu
 275 280 285

Glu Ile Phe Ala Arg Gln Leu Asp Leu Leu Thr Leu Glu Asn Ile Lys
 290 295 300

Gln Thr Glu Glu Ala Leu Asp Asn His Arg Leu Pro Phe Pro Leu Leu
 305 310 315 320

Asp Ala Gly Thr Ile Lys
 325

<210> 39

<211> 320

<212> PRT

<213> Artificial Sequence

<220>

<223> GUSn/Intn(6) fusion, with 6 amino acid deletion in Intn

<220>

<221> PEPTIDE

<222> (1)..(320)

<223> GUSn/Intn(6) fusion, with 6 amino acid deletion in Intn

<400> 39

Met Ala His His His His His Val Arg Pro Val Glu Thr Pro Thr
 1 5 10 15

Arg Glu Ile Lys Lys Leu Asp Gly Leu Trp Ala Phe Ser Leu Asp Arg
 20 25 30

Glu Asn Cys Gly Ile Asp Gln Arg Trp Trp Glu Ser Ala Leu Gln Glu
 35 40 45

Ser Arg Ala Ile Ala Val Pro Gly Ser Phe Asn Asp Gln Phe Ala Asp
 50 55 60

Ala Asp Ile Arg Asn Tyr Ala Gly Asn Val Trp Tyr Gln Arg Glu Val
 65 70 75 80

Phe Ile Pro Lys Gly Trp Ala Gly Gln Arg Ile Val Leu Arg Phe Asp
 85 90 95

Ala Val Thr His Tyr Gly Lys Val Trp Val Asn Asn Gln Glu Val Met
 100 105 110

Glu His Gln Gly Gly Tyr Thr Pro Phe Glu Ala Asp Val Thr Pro Tyr
 115 120 125

Val Ile Ala Gly Lys Ser Val Arg Ile Thr Val Cys Val Asn Asn Glu
 Page 14

130

135

140

Leu Asn Trp Gln Thr Ile Pro Pro Gly Met Val Ile Thr Asp Glu Asn
 145 150 155 160

Gly Lys Lys Lys Gln Ser Tyr Phe His Asp Phe Phe Asn Tyr Ala Gly
 165 170 175

Ile His Arg Ser Val Met Leu Tyr Thr Thr Pro Asn Thr Trp Val Asp
 180 185 190

Asp Ile Thr Val Val Thr His Val Ala Gln Asp Glu Ile Leu Thr Val
 195 200 205

Glu Tyr Gly Pro Leu Pro Ile Gly Lys Ile Val Ser Glu Glu Ile Asn
 210 215 220

Cys Ser Val Tyr Ser Val Asp Pro Glu Gly Arg Val Tyr Thr Gln Ala
 225 230 235 240

Ile Ala Gln Trp His Asp Arg Gly Glu Gln Glu Val Leu Glu Tyr Glu
 245 250 255

Leu Glu Asp Gly Ser Val Ile Arg Ala Thr Ser Asp His Arg Phe Leu
 260 265 270

Thr Thr Asp Tyr Gln Leu Leu Ala Ile Glu Glu Ile Phe Ala Arg Gln
 275 280 285

Leu Asp Leu Leu Thr Leu Glu Asn Ile Lys Gln Thr Glu Glu Ala Leu
 290 295 300

Asp Asn His Arg Leu Pro Phe Pro Leu Leu Asp Ala Gly Thr Ile Lys
 305 310 315 320

<210> 40
 <211> 450
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Intc/GUSc fusion

<220>
 <221> PEPTIDE
 <222> (1)..(450)
 <223> Intc/GUSc fusion

<400> 40

Met Val Lys Val Ile Gly Arg Arg Ser Leu Gly Val Gln Arg Ile Phe
 1 5 10 15

Asp Ile Gly Leu Pro Gln Asp His Asn Phe Leu Leu Ala Asn Gly Ala
 Page 15

20

25

30

Ile Ala Ala Asn Cys Asn His Ala Ser Val Asp Trp Gln Val Val Ala
 35 40 45

Asn Gly Asp Val Ser Val Glu Leu Arg Asp Ala Asp Gln Gln Val Val
 50 55 60

Ala Thr Gly Gln Gly Thr Ser Gly Thr Leu Gln Val Val Asn Pro His
 65 70 75 80

Leu Trp Gln Pro Gly Glu Gly Tyr Leu Tyr Glu Leu Cys Val Thr Ala
 85 90 95

Lys Ser Gln Thr Glu Cys Asp Ile Tyr Pro Leu Arg Val Gly Ile Arg
 100 105 110

Ser Val Ala Val Lys Gly Glu Gln Phe Leu Ile Asn His Lys Pro Phe
 115 120 125

Tyr Phe Thr Gly Phe Gly Arg His Glu Asp Ala Asp Leu Arg Gly Lys
 130 135 140

Gly Phe Asp Asn Val Leu Met Val His Asp His Ala Leu Met Asp Trp
 145 150 155 160

Ile Gly Ala Asn Ser Tyr Arg Thr Ser His Tyr Pro Tyr Ala Glu Glu
 165 170 175

Met Leu Asp Trp Ala Asp Glu His Gly Ile Val Val Ile Asp Glu Thr
 180 185 190

Ala Ala Val Gly Phe Asn Leu Ser Leu Gly Ile Gly Phe Glu Ala Gly
 195 200 205

Asn Lys Pro Lys Glu Leu Tyr Ser Glu Glu Ala Val Asn Gly Glu Thr
 210 215 220

Gln Gln Ala His Leu Gln Ala Ile Lys Glu Leu Ile Ala Arg Asp Lys
 225 230 235 240

Asn His Pro Ser Val Val Met Trp Ser Ile Ala Asn Glu Pro Asp Thr
 245 250 255

Arg Pro Gln Gly Ala Arg Glu Tyr Phe Ala Pro Leu Ala Glu Ala Thr
 260 265 270

Arg Lys Leu Asp Pro Thr Arg Pro Ile Thr Cys Val Asn Val Met Phe
 275 280 285

Cys Asp Ala His Thr Asp Thr Ile Ser Asp Leu Phe Asp Val Leu Cys
 Page 16

290

295

300

Leu Asn Arg Tyr Tyr Gly Trp Tyr Val Gln Ser Gly Asp Leu Glu Thr
 305 310 315 320

Ala Glu Lys Val Leu Glu Lys Glu Leu Leu Trp Gln Glu Lys Leu His
 325 330 335

Gln Pro Ile Ile Thr Glu Tyr Gly Val Asp Thr Leu Ala Gly Leu
 340 345 350

His Ser Met Tyr Thr Asp Met Trp Ser Glu Glu Tyr Gln Cys Ala Trp
 355 360 365

Leu Asp Met Tyr His Arg Val Phe Asp Arg Val Ser Ala Val Val Gly
 370 375 380

Glu Gln Val Trp Asn Phe Ala Asp Phe Ala Thr Ser Gln Gly Ile Leu
 385 390 395 400

Arg Val Gly Gly Asn Lys Lys Gly Ile Phe Thr Arg Asp Arg Lys Pro
 405 410 415

Lys Ser Ala Ala Phe Leu Leu Gln Lys Arg Trp Thr Gly Met Asn Phe
 420 425 430

Gly Glu Lys Pro Gln Gln Gly Gly Lys Gln Gly Ser His His His His
 435 440 445

His His
 450

<210> 41
 <211> 41
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer to amplify NOS terminator region

<400> 41
 gcgtcgacag tcactctaga gacatcgatc tagtaacata g

41

<210> 42
 <211> 41
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer to amplify NOS terminator region

<400> 42
 ggggtacccc atgcggccgc ctaaagaagg agtgcgtcga a

41

<210> 43

CL-1806-conversion-CIP 3-5-04.ST25

<211> 18
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> 18 bp polylinker

 <400> 43
 ggtacccgat ccaattcc 18

 <210> 44
 <211> 26
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Primer

 <400> 44
 gaccatggcc aatttactga ccgtac 26

 <210> 45
 <211> 27
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Primer

 <400> 45
 cgaaagaaag gcagcagcga tcgctat 27

 <210> 46
 <211> 29
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Primer

 <400> 46
 atagcgatcg ctgctgcctt tctttcgga 29

 <210> 47
 <211> 27
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Primer

 <400> 47
 atgtcgactc acttgatggt tccagca 27

 <210> 48
 <211> 3034
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Sequence of 3034 bp Asp 718 fragment containing
 35S-CreN-IntN-3'ocs gene in plasmid pGV947

CL-1806-conversion-CIP 3-5-04.ST25

<400> 48
 ggtacccgat ccaattccaa tcccacaaaa atctgagctt aacagcacag ttgctcctct 60
 cagagcagaa tcgggtattc aacaccctca tatcaactac tacgttgtgt ataacggtcc 120
 acatgccggt atatacgatg actgggggttg tacaaaggcg gcaacaaacg gcgttcccgg 180
 agttgcacac aagaaatttg ccactattac agaggcaaga gcagcagctg acgcgtacac 240
 aacaagtcag caaacagaca ggttgaactt catcccaaaa ggagaagctc aactcaagcc 300
 caagagcttt gctaaggccc taacaagccc accaaagcaa aaagcccact ggctcacgct 360
 aggaacaaaa agggccagca gtgatccagc cccaaaagag atctcctttg ccccgagat 420
 tacaatggac gatttcctct atctttacga tctaggaagg aagttcgaag gtgaaggtga 480
 cgacactatg ttcaccactg ataatgagaa ggttagcctc ttcaatttca gaaagaatgc 540
 tgaccacag atggttagag aggcctacgc agcaggctctc atcaagacga tctacccgag 600
 taacaatctc caggagatca aataccttcc caagaagggt aaagatgcag tcaaaagatt 660
 caggactaat tgcatacaaga acacagagaa agacatatct ctcaagatca gaagtactat 720
 tccagtatgg acgattcaag gcttgcttca taaaccaagg caagtaatag agattggagt 780
 ctctaaaaag gtagttccta ctgaatctaa ggccatgcat ggagtctaag attcaaactg 840
 aggatctaac agaactcgcc gtgaagactg gcgaacagtt catacagagt cttttacgac 900
 tcaatgacaa gaagaaaatc ttcgtcaaca tgggtggagca cgacactctg gtctactcca 960
 aaaatgtcaa agatacagtc tcagaagacc aaagggtctat tgagactttt caacaaagga 1020
 taatttcggg aaacctcctc ggattccatt gccagctat ctgtcacttc atcgaaagga 1080
 cagtagaaaa ggaagggtggc tcctacaaat gccatcattg cgataaagga aaggctatca 1140
 ttcaagatgc ctctgccgac agtgggtcca aagatggacc cccaccacg aggagcatcg 1200
 tggaaaaaga agacgttcca accacgtctt caaagcaagt ggattgatgt gacatctcca 1260
 ctgacgtaag ggatgacgca caatcccact atccttcgca agacccttcc tctatataag 1320
 gaagttcatt tcatttggag aggacacgct cgagctcatt tctctattac ttcagccata 1380
 acaaaagaac tcttttctct tcttattaaa ccatggccaa tttactgacc gtacacaaaa 1440
 atttgcctgc attaccggtc gatgcaacga gtgatgaggt tcgcaagaac ctgatggaca 1500
 tgttcagggg tcgccaggcg ttttctgagc atacctggaa aatgcttctg tccgtttgcc 1560
 ggtcgtgggc ggcatggtgc aagttgaata accggaaatg gtttcccgca gaacctgaag 1620
 atgttcgca ttatcttcta tatcttcagg cgcgcggtct ggagtaaaa actatccagc 1680
 aacatttggg ccagctaaac atgcttcacg gtcggtccgg gctgccacga ccaagtgaac 1740
 gcaatgctgt ttcactagtt atgcggcgga tccgaaaaga aaacgttgat gccggtgaac 1800
 gtgcaaaaca ggctctagcg ttcgaacgca ctgatttcga ccaggttcgt tcactcatgg 1860
 aaaatagcga tcgctgctgc ctttctttcg gaactgagat cttaccgtt gagtacggac 1920
 cacttcctat tggttaagatc gtttctgagg aaattaactg ctcaagtgtac tctgttgatc 1980
 cagaaggaag agtttacact caggctatcg cacaatggca cgataggggt gaacaagagg 2040

CL-1806-conversion-CIP 3-5-04.ST25

| | |
|---|------|
| ttctcgagta cgagcttgaa gatggatccg ttattcgtgc tacctctgac catagattct | 2100 |
| tgactacaga ttatcagctt ctcgctatcg aggaaatctt tgctaggcaa cttgatctcc | 2160 |
| ttactttgga gaacatcaag cagacagaag aggctcttga caaccacaga cttccattcc | 2220 |
| ctttgctcga tgctggaacc atcaagtga tgcacataat cactagagtc ctgctttaat | 2280 |
| gagatatgcg agacgcctat gatcgcataa tatttgcttt caattctggt gtgcacgttg | 2340 |
| taaaaaacct gagcatgtgt agctcagatc cttaccgccg gtttcggttc attctaata | 2400 |
| atatatcacc cgttactatc gtatttttat gaataatatt ctccgttcaa tttactgatt | 2460 |
| gtaccctact acttatatgt acaatattaa aatgaaaaca atatattgtg ctgaataggt | 2520 |
| ttatagcgac atctatgata gagcgccaca ataacaaaca attgcgtttt attattacaa | 2580 |
| atccaatttt aaaaaaagcg gcagaaccgg tcaaacctaa aagactgatt acataaatct | 2640 |
| tattcaaatt tcaaaaggcc ccaggggcta gtatctacga cacaccgagc ggcgaactaa | 2700 |
| taacgttcac tgaagggaac tccggttccc cgccggcgcg catgggtgag attccttgaa | 2760 |
| gttgagtatt ggccgtccgc tctaccgaaa gttacgggca ccattcaacc cgggccagca | 2820 |
| cggcgccgg gtaaccgact tgctgccccg agaattatgc agcatttttt tgggtgatgt | 2880 |
| gggccccaaa tgaagtgcag gtcaaacctt gacagtgcag acaaatcggt gggcgggtcc | 2940 |
| agggcgaatt ttgcgacaac atgtcgaggc tcagcaggac ctgcaggcat gcaagcttat | 3000 |
| cgataccgctc gacctcgagg gggggccccg tacc | 3034 |

<210> 49
 <211> 17
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> 17 bp linker sequence

<400> 49
 gtcgacataa tcactag 17

<210> 50
 <211> 60
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> 60 bp polylinker

<400> 50
 caggacctgc aggcatgcaa gcttatcgat accgtcgacc tcgagggggg gcccgggtacc 60

<210> 51
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer

<400> 51
gaccatgggtt aagggtgattg gaagacg 27

<210> 52
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 52
tacgtatatc ctggcaattg gcagcgatgg 30

<210> 53
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 53
cgctgccaat tgccaggata tacgtaatct g 31

<210> 54
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 54
agtcgaccta atcgccatct tccagcag 28

<210> 55
<211> 2873
<212> DNA
<213> Artificial Sequence

<220>
<223> Sequence of 2873 bp Asp718 fragment containing
35S:IntC-CreC:3'ocs in plasmid pGV951

<400> 55
ggtacccgat ccaattccaa tcccacaaaa atctgagctt aacagcacag ttgctcctct 60
cagagcagaa tcgggtattc aacaccctca tatcaactac tacgttgtgt ataacgggtcc 120
acatgccggt atatacgatg actgggggttg taaaaaggcg gcaacaaacg gcgttcccgg 180
agttgcacac aagaaatttg ccactattac agaggcaaga gcagcagctg acgcgtacac 240
aacaagtcag caaacagaca ggttgaactt catccccaaa ggagaagctc aactcaagcc 300
caagagcttt gctaaggccc taacaagccc accaaagcaa aaagcccact ggctcacgct 360
aggaaccaaa aggcccagca gtgatccagc cccaaaagag atctcctttg ccccgagat 420
tacaatggac gatttcctct atctttacga tctaggaagg aagttcgaag gtgaagggtga 480
cgacactatg ttcaccactg ataatgagaa ggtagcctc ttcaatttca gaaagaatgc 540

CL-1806-conversion-CIP 3-5-04.ST25

| | | | | | | |
|-------------|-------------|-------------|-------------|------------|------------|------|
| tgacccacag | atgggttagag | aggcctacgc | agcagggtctc | atcaagacga | tctacccgag | 600 |
| taacaatctc | caggagatca | aataccttcc | caagaagggtt | aaagatgcag | tcaaaagatt | 660 |
| caggactaat | tgcatcaaga | acacagagaa | agacatatatt | ctcaagatca | gaagtactat | 720 |
| tccagtatgg | acgattcaag | gcttgcttca | taaaccaagg | caagtaatag | agattggagt | 780 |
| ctctaaaaag | gtagttccta | ctgaatctaa | ggccatgcat | ggagtctaag | attcaaatcg | 840 |
| aggatctaac | agaactcgcc | gtgaagactg | gcgaacagtt | catacagagt | cttttacgac | 900 |
| tcaatgacaa | gaagaaaatc | ttcgtcaaca | tggtggagca | cgacactctg | gtctactcca | 960 |
| aaaatgtcaa | agatacagtc | tcagaagacc | aaagggctat | tgagactttt | caacaaagga | 1020 |
| taatttcggg | aaacctcctc | ggattccatt | gcccagctat | ctgtcacttc | atcgaaagga | 1080 |
| cagtagaaaa | ggaagggtggc | tcctacaaat | gccatcattg | cgataaagga | aaggctatca | 1140 |
| ttcaagatgc | ctctgccgac | agtgggtcca | aagatggacc | cccaccacg | aggagcatcg | 1200 |
| tggaaaaaaga | agacgttcca | accacgtctt | caaagcaagt | ggattgatgt | gacatctcca | 1260 |
| ctgacgtaag | ggatgacgca | caatcccact | atccttcgca | agacccttcc | tctatataag | 1320 |
| gaagttcatt | tcatttgag | aggacacgct | cgagctcatt | tctctattac | ttcagccata | 1380 |
| acaaaagaac | tcttttctct | tcttattaaa | ccatgggttaa | ggtgattgga | agacgttctc | 1440 |
| ttggtgttca | aaggatcttc | gatatcggat | tgccacaaga | ccacaacttt | cttctcgcta | 1500 |
| atggtgccat | cgctgccaat | tgccaggata | tacgtaatct | ggcatttctg | gggattgctt | 1560 |
| ataacaccct | gttacgtata | gccgaaattg | ccaggatcag | ggttaaagat | atctcacgta | 1620 |
| ctgacggtgg | gagaatgtta | atccatattg | gcagaacgaa | aacgctgggt | agcaccgcag | 1680 |
| gtgtagagaa | ggcacttagc | ctgggggttaa | ctaaactggt | cgagcgatgg | atttccgtct | 1740 |
| ctggtgtagc | tgatgatccg | aataactacc | tgttttgccg | ggtcagaaaa | aatggtgttg | 1800 |
| ccgcgccatc | tgccaccagc | cagctatcaa | ctcgcgccct | ggaagggatt | tttgaagcaa | 1860 |
| ctcatcgatt | gatttacggc | gctaaggatg | actctggtca | gagatacttg | gcctggtctg | 1920 |
| gacacagtg | ccgtgtcgga | gccgcgcgag | atatggcccc | cgctggagtt | tcaataccgg | 1980 |
| agatcatgca | agctggtggc | tggaccaatg | taaatattgt | catgaactat | atccgtaacc | 2040 |
| tggatagtga | aacaggggca | atggtgcgcc | tgctggaaga | tggcgattag | gtcgactatc | 2100 |
| actagagtcc | tgctttaatg | agatatgcga | gacgcctatg | atcgcatgat | atttgctttc | 2160 |
| aattctgttg | tgcacgttgt | aaaaaacctg | agcatgtgta | gctcagatcc | ttaccgccgg | 2220 |
| tttcggttca | ttctaataaa | tatatcacc | gttactatcg | tatttttatg | aataatattc | 2280 |
| tccgttcaat | ttactgattg | taccctacta | cttatatgta | caatattaaa | atgaaaacaa | 2340 |
| tatatgtg | tgaatagggt | tatagcgaca | tctatgatag | agcgccacaa | taacaaacaa | 2400 |
| ttgcgtttta | ttattacaaa | tccaatttta | aaaaaagcgg | cagaaccggt | caaacctaaa | 2460 |
| agactgatta | cataaatctt | attcaaattt | caaaaggccc | caggggctag | tatctacgac | 2520 |
| acaccgagcg | gcgaactaat | aacgttcact | gaagggaact | ccggttcccc | gccggcgcg | 2580 |

CL-1806-conversion-CIP 3-5-04.ST25

| | | | | | | |
|------------|------------|------------|------------|------------|------------|------|
| atgggtgaga | ttccttgaag | ttgagtattg | gccgtccgct | ctaccgaaag | ttacgggcac | 2640 |
| cattcaaccc | ggtccagcac | ggcggccggg | taaccgactt | gctgccccga | gaattatgca | 2700 |
| gcattttttt | ggtgtatgtg | ggcccccatt | gaagtgcagg | tcaaaccttg | acagtgcgca | 2760 |
| caaatcgttg | ggcgggtcca | gggcgaattt | tgcgacaaca | tgctgagggt | cagcaggacc | 2820 |
| tgaggcatg | caagcttatc | gataccgtcg | acctcgaggg | ggggcccggg | acc | 2873 |

<210> 56
 <211> 14
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> 15 bp linker

| | |
|------------|------|
| <400> 56 | |
| tcgactatca | ctag |
| | 14 |

<210> 57
 <211> 5449
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Sequence of 5449 bp Sal I-HindIII fragment containing the blocked GUS reporter gene for Cre-Lox excision in plasmid pGV801

| | | | | | | |
|-------------|------------|------------|-------------|-------------|-------------|------|
| <400> 57 | | | | | | |
| gtcgactcta | gaggatccaa | ttccaatccc | acaaaaatct | gagcttaaca | gcacagttgc | 60 |
| tcctctcaga | gcagaatcgg | gtattcaaca | ccctcatatc | aactactacg | ttgtgtataa | 120 |
| cgggtccacat | gccggtatat | acgatgactg | gggttggtaca | aaggcggcaa | caaacggcgt | 180 |
| tcccggagtt | gcacacaaga | aatttgccac | tattacagag | gcaagagcag | cagctgacgc | 240 |
| gtacacaaca | agtcagcaaa | cagacagggt | gaacttcatc | cccaaaggag | aagctcaact | 300 |
| caagcccaag | agctttgcta | aggccctaac | aagcccacca | aagcaaaaag | cccactggct | 360 |
| cacgctagga | acaaaaaggc | ccagcagtga | tccagcccca | aaagagatct | cctttgcccc | 420 |
| ggagattaca | atggacgatt | tcctctatct | ttacgatcta | ggaaggaagt | tcgaagggtga | 480 |
| aggtgacgac | actatgttca | ccactgataa | tgagaagggt | agcctcttca | atttcagaaa | 540 |
| gaatgctgac | ccacagatgg | ttagagaggc | ctacgcagca | ggtctcatca | agacgatcta | 600 |
| cccagagtaac | aatctccagg | agatcaaata | ccttcccaag | aagggttaaag | atgcagtcaa | 660 |
| aagattcagg | actaattgca | tcaagaacac | agagaaagac | atatttctca | agatcagaag | 720 |
| tactattcca | gtatggacga | ttcaaggctt | gcttcataaa | ccaaggcaag | taatagagat | 780 |
| tggagtctct | aaaaaggtag | ttcctactga | atctaaggcc | atgcatggag | tctaagattc | 840 |
| aaatcgagga | tctaacagaa | ctcgccgtga | agactggcga | acagttcata | cagagtcttt | 900 |
| tacgactcaa | tgacaagaag | aaaatcttcg | tcaacatggg | ggagcacgac | actctgggtct | 960 |
| actccaaaaa | tgtcaaagat | acagtctcag | aagaccaaag | ggctattgag | acttttcaac | 1020 |
| aaaggataat | ttcgggaaac | ctcctcggat | tccattgccc | agctatctgt | cacttcatcg | 1080 |

CL-1806-conversion-CIP 3-5-04.ST25

| | | | | | | |
|------------|-------------|-------------|------------|-------------|------------|------|
| aaaggacagt | agaaaaggaa | ggtggctcct | acaaatgcc | tcattgcat | aaaggaaagg | 1140 |
| ctatcattca | agatgcctct | gccgacagt | gtcccaaaga | tggaccccca | cccacgagga | 1200 |
| gcatcgtgga | aaaagaagac | gttccaacca | cgtcttcaaa | gcaagtggat | tgatgtgaca | 1260 |
| tctccactga | cgtaagggat | gacgcacaat | cccactatcc | ttcgcaagac | ccttcctcta | 1320 |
| tataaggaag | ttcatttcat | ttggagagga | cacgctcgag | ctcatttctc | tattacttca | 1380 |
| gccataacaa | aagaactctt | ttctcttctt | attaaaccat | gataacttcg | tatagcatac | 1440 |
| attatacgaa | gttatcctag | gatcatgagc | ggagaattaa | gggagtcacg | ttatgacccc | 1500 |
| cgccgatgac | gcgggacaag | ccgttttacg | tttggaactg | acagaaccgc | aacgttgaag | 1560 |
| gagccactca | gccgcgggtt | tctggagttt | aatgagctaa | gcacatacgt | cagaaaccat | 1620 |
| tattgcgctg | tcaaaagtcg | cctaagggtc | ctatcagcta | gcaaataattt | cttgtcaaaa | 1680 |
| atgctccact | gacgttccat | aaattcccct | cggtatccaa | ttagagtctc | atattcactc | 1740 |
| tcaatccaaa | taatctgcac | cggatctgga | tcgtttcgca | tgattgaaca | agatggattg | 1800 |
| cacgcagggt | ctccggccgc | ttgggtggag | aggctattcg | gctatgactg | ggcacaacag | 1860 |
| acaatcggct | gctctgatgc | cgccgtgttc | cggtgtcag | cgaggggcg | cccggttctt | 1920 |
| tttgtcaaga | ccgacctgtc | cggtgccctg | aatgaactgc | aggacgaggc | agcgcggcta | 1980 |
| tcgtggctgg | ccacgacggg | cgttccttgc | gcagctgtgc | tcgacgttgt | cactgaagcg | 2040 |
| ggaagggact | ggctgctatt | gggcgaagtg | ccggggcagg | atctcctgtc | atctcacctt | 2100 |
| gctcctgccg | agaaagtatc | catcatggct | gatgcaatgc | ggcggctgca | tacgcttgat | 2160 |
| ccggctacct | gccattcgca | ccaccaagcg | aaacatcgca | tcgagcgagc | acgtactcgg | 2220 |
| atggaagccg | gtcttgctga | tcaggatgat | ctggacgaag | agcatcaggg | gctcgcgcca | 2280 |
| gccgaactgt | tcgccaggct | caaggcgcg | atgcccgacg | gcgatgatct | cgtcgtgacc | 2340 |
| catggcgatg | cctgcttgcc | gaatatcatg | gtggaaaatg | gccgcttttc | tggattcatc | 2400 |
| gactgtggcc | ggctgggtgt | ggcggaccgc | tatcaggaca | tagcgttggc | taccctgat | 2460 |
| attgctgaag | agcttggcgg | cgaatgggct | gaccgcttcc | tcgtgcttta | cggtatcgcc | 2520 |
| gctcccgaat | cgcagcgcat | cgccttctat | cgccttcttg | acgagttctt | ctgagcggga | 2580 |
| ctctgggggt | cgaaatgacc | gaccaagcga | cgcccaacct | gccatcacga | gatttcgatt | 2640 |
| ccaccgccgc | cttctatgaa | aggttgggct | tcggaatcgt | tttccgggac | gccggctgga | 2700 |
| tgatcctcca | gcgcggggat | ctcatgctgg | agttcttcgc | ccacgggatc | tctgcggaac | 2760 |
| aggcggtcga | aggtgccgat | atcattacga | cagcaacggc | cgacaagcac | aacgccacga | 2820 |
| tcctgagcga | caatatgatc | gggcccggcg | tccacatcaa | cggcgtcggc | ggcgactgcc | 2880 |
| caggcaagac | cgagatgcac | cgcgatatct | tgctgcgttc | ggatattttc | gtggagtctc | 2940 |
| cgccacagac | ccggatgatc | cccgatcggt | caaacatttg | gcaataaagt | ttcttaagat | 3000 |
| tgaatcctgt | tgccgggtctt | gcgatgatta | tcatataatt | tctgttgaat | tacgttaagc | 3060 |
| atgtaataat | taacatgtaa | tgcgatgacgt | tatttatgag | atgggttttt | atgattagag | 3120 |

CL-1806-conversion-CIP 3-5-04.ST25

| | | | | | | |
|-------------|------------|-------------|-------------|------------|------------|------|
| tcccgcgaatt | atacatttaa | tacgcgatag | aaaacaaaat | atagcgcgca | aactaggata | 3180 |
| aattatcgcg | cgcggtgtca | tctatgttac | tagatcgggc | ctcctgtcaa | tgctggccta | 3240 |
| ggtaaataac | ttcgtatagc | atacattata | cgaagttatt | agtacgtcct | gtagaaaacc | 3300 |
| caacccgtga | aatcaaaaaa | ctcgacggcc | tgtgggcatt | cagtctggat | cgcgaaaact | 3360 |
| gtggaattga | tcagcgttgg | tgggaaagcg | cgttacaaga | aagccgggca | attgctgtgc | 3420 |
| caggcagttt | taacgatcag | ttcgccgatg | cagatattcg | taattatgcg | ggcaacgtct | 3480 |
| ggtatcagcg | cgaagtcttt | ataccgaaag | gttgggcagg | ccagcgtatc | gtgctgcggt | 3540 |
| tcgatgcggt | cactcattac | ggcaaagtgt | gggtcaataa | tcaggaagtg | atggagcatc | 3600 |
| agggcggcta | tacgccattt | gaagccgatg | tcacgccgta | tgttattgcc | gggaaaagtg | 3660 |
| tacgtatcac | cgtttgtgtg | aacaacgaac | tgaactggca | gactatcccg | ccgggaatgg | 3720 |
| tgattaccga | cgaaaacggc | aagaaaaagc | agtcttactt | ccatgatttc | tttaactatg | 3780 |
| ccggaatcca | tcgcagcgta | atgctctaca | ccacgccgaa | cacctgggtg | gacgatatca | 3840 |
| ccgtggtgac | gcatgtcgcg | caagactgta | accacgcgtc | tgttgactgg | caggtggtgg | 3900 |
| ccaatggtga | tgtcagcggt | gaactgcgtg | atgcggatca | acaggtgggt | gcaactggac | 3960 |
| aaggcactag | cgggactttg | caagtgggtga | atccgcacct | ctggcaaccg | ggtgaagggt | 4020 |
| atctctatga | actgtgcgtc | acagccaaaa | gccagacaga | gtgtgatatc | tacccgcttc | 4080 |
| gcgtcggcat | ccggtcagtg | gcagtgaagg | gccaacagtt | cctgattaac | cacaaaccgt | 4140 |
| tctactttac | tggctttggg | cgatcatgaag | atgcggactt | acgtggcaaa | ggattcgata | 4200 |
| acgtgctgat | ggtgcacgac | cacgcattaa | tggactggat | tggggccaac | tcctaccgta | 4260 |
| cctcgcatta | cccttacgct | gaagagatgc | tcgactgggc | agatgaacat | ggcatcgtgg | 4320 |
| tgattgatga | aactgctgct | gtcggcttta | acctctcttt | aggcattggg | ttcgaagcgg | 4380 |
| gcaacaagcc | gaaagaactg | tacagcgaag | aggcagtcaa | cggggaaact | cagcaagcgc | 4440 |
| acttacaggc | gattaaagag | ctgatagcgc | gtgacaaaaa | ccaccaagc | gtggtgatgt | 4500 |
| ggagtattgc | caacgaaccg | gatacccgtc | cgcaagtgca | cgggaatatt | tcgccactgg | 4560 |
| cggaaagcaac | gcgtaaactc | gacccgacgc | gtccgatcac | ctgcgtcaat | gtaatgttct | 4620 |
| gcgacgctca | caccgatacc | atcagcgatc | tctttgatgt | gctgtgcctg | aaccgttatt | 4680 |
| acggatggta | tgtccaaagc | ggcgatttgg | aaacggcaga | gaaggtagtg | gaaaaagaac | 4740 |
| ttctggcctg | gcaggagaaa | ctgcatcagc | cgattatcat | caccgaatac | ggcgtggata | 4800 |
| cgtagccgg | gctgcactca | atgtacaccg | acatgtggag | tgaagagtat | cagtgtgcat | 4860 |
| ggctggatat | gtatcaccgc | gtctttgatc | gcgtcagcgc | cgtcgtcggg | gaacaggtat | 4920 |
| ggaatttcgc | cgattttgcg | acctcgcaag | gcatattgcg | cgttggcggg | aacaagaaag | 4980 |
| ggatcttcac | tcgcgaccgc | aaaccgaagt | cggcggcttt | tctgctgcaa | aaacgctgga | 5040 |
| ctggcatgaa | cttcggtgaa | aaaccgcagc | agggaggcaa | acaatgaatc | aacaactctc | 5100 |
| ctggcgcacc | atcgtcgggt | acagcctcgg | tgggggaattc | cccgggggta | cctaaagaag | 5160 |

CL-1806-conversion-CIP 3-5-04.ST25

gagtgcgtcg aagcagatcg ttcaaacatt tggcaataaa gtttcttaag attgaatcct 5220
 gttgccgggtc ttgcgatgat tatcatataa tttctgttga attacgttaa gcatgtaata 5280
 attaacaatgt aatgcatgac gttatattatg agatggggtt ttatgattag agtcccgcaa 5340
 ttatacatatt aatacgcgat agaaaacaaa atatagcgcg caaactagga taaattatcg 5400
 cgcgcggtgt catctatggt actagatcga tgtcgactct agaaagctt 5449

<210> 58
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> 24 bp polylinker

<400> 58
 gtcgactcta gaggatccaa ttcc 24

<210> 59
 <211> 34
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Lox P sequence

<400> 59
 ataacttcgt atagcataca ttatacgaag ttat 34

<210> 60
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> 22 bp polylinker

<400> 60
 tggggaattc cccgggggta cc 22

<210> 61
 <211> 18
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> 18 bp polylinker

<400> 61
 gtcgactcta gaaagctt 18

<210> 62
 <211> 605
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Elastin-based protein polymer (Zhang et al., Plant Cell Rep. 16(3-4):174-179 (1996))

<400> 62

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
1 5 10 15

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
20 25 30

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly
35 40 45

Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val
50 55 60

Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
65 70 75 80

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
85 90 95

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
100 105 110

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly
115 120 125

Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val
130 135 140

Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
145 150 155 160

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
165 170 175

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
180 185 190

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly
195 200 205

Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val
210 215 220

Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
225 230 235 240

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
245 250 255

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
260 265 270

260

265

270

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly
275 280 285

Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val
290 295 300

Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
305 310 315 320

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
325 330 335

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
340 345 350

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly
355 360 365

Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val
370 375 380

Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
385 390 395 400

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
405 410 415

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
420 425 430

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly
435 440 445

Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val
450 455 460

Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
465 470 475 480

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
485 490 495

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
500 505 510

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly
515 520 525

Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val
530 535 540 545

530

Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
545 550 555 560

Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly
565 570 575

Val Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro Gly Val
580 585 590

Gly Val Pro Gly Val Gly Val Pro Gly Val Gly Val Pro
595 600 605

<210> 63
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Coding sequence introduced by oligomer HGUSH-n
<400> 63

Met Ala His His His His His His
1 5

<210> 64
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Insertion sequence in pGY101 (a pBluscript-based plasmid)
<400> 64

Met Ala Arg Ser Arg Gly Ser His His His His His His
1 5 10

<210> 65
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Residues deleted from IntN to create GUSn/Intn(6) fusion
<400> 65

Cys Leu Ser Phe Gly Thr
1 5

<210> 66
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> 12-amino acid N-terminal amino acid extension to GUS ORF
Page 29

<400> 66

Ile Thr Ser Tyr Ser Ile His Tyr Thr Lys Leu Leu
 1 5 10

<210> 67

<211> 21

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 67

gtatcaccgt ttgtgtgaac a

21

<210> 68

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 68

ggaattcctc agatgttctc caaagtaagg ag

32

<210> 69

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 69

ggaattcctc aatctgtagt caagaatcta tg

32

<210> 70

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> primer

<400> 70

ggaattcctc aaacctcttg ttcaccccta tc

32

<210> 71

<211> 78

<212> PRT

<213> Synechocystis sp. PCC6803

<400> 71

Cys Leu Ser Phe Gly Thr Glu Ile Leu Thr Val Glu Tyr Gly Pro Leu
 1 5 10 15

Pro Ile Gly Lys Ile Val Ser Glu Glu Ile Asn Cys Ser Val Tyr Ser
 20 25 30

Val Asp Pro Glu Gly Arg Val Tyr Thr Gln Ala Ile Ala Gln Trp His
 35 40 45

Asp Arg Gly Glu Gln Glu Val Leu Glu Tyr Glu Leu Glu Asp Gly Ser
 50 55 60

Val Ile Arg Ala Thr Ser Asp His Arg Phe Leu Thr Thr Asp
 65 70 75

<210> 72
 <211> 234
 <212> DNA
 <213> artificial sequence

<220>
 <223> plant codon optimized IntN encoding sequence

<400> 72
 tgcctttctt tcggaactga gatccttacc gttgagtacg gaccacttcc tattggtaag 60
 atcgtttctg aggaaattaa ctgctcagtg tactctgttg atccagaagg aagagtttac 120
 actcaggcta tcgcacaatg gcacgatagg ggtgaacaag aggttctcga gtacgagctt 180
 gaagatggat ccggttattcg tgctacctct gaccatagat tcttgactac agat 234

<210> 73
 <211> 12
 <212> PRT
 <213> Synechocystis sp. PCC6803

<400> 73

Cys Leu Ser Phe Gly Thr Glu Ile Leu Thr Val Glu
 1 5 10

<210> 74
 <211> 15
 <212> PRT
 <213> Synechocystis sp. PCC6803

<400> 74

Asp Gly Ser Val Ile Arg Ala Thr Ser Asp His Arg Phe Leu Thr
 1 5 10 15

<210> 75
 <211> 13
 <212> PRT
 <213> Trichodesmium erythraeum

<400> 75

Cys Leu Thr Tyr Glu Thr Glu Ile Met Thr Val Glu Tyr
 1 5 10

<210> 76
 <211> 15
 <212> PRT

<213> Trichodesmium erythraeum

<400> 76

Asp Gly Thr Val Ile Arg Ala Thr Pro Glu His Lys Phe Met Thr
1 5 10 15

<210> 77

<211> 12

<212> PRT

<213> artificial

<220>

<223> compiled sequence

<220>

<221> misc_feature

<222> (3)..(5)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (9)..(9)

<223> Xaa can be any naturally occurring amino acid

<400> 77

Cys Leu Xaa Xaa Xaa Thr Glu Ile Xaa Thr Val Glu
1 5 10

<210> 78

<211> 15

<212> PRT

<213> artificial

<220>

<223> compiled sequence

<220>

<221> misc_feature

<222> (3)..(3)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (9)..(10)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (12)..(12)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (14)..(14)

<223> Xaa can be any naturally occurring amino acid

<400> 78

Asp Gly Xaa Val Ile Arg Ala Thr Xaa Xaa His Xaa Phe Xaa Thr
1 5 10 15